语义信息分解方法

石光明[1,2,3]，梁宇平[1]，高大化[2,3]，杨旻曦[1]，李海雄[1]，宋晓丹[1]

[1]（西安电子科技大学 西安 710071）

[2]（鹏城实验室 深圳 518000）

[3]（琶洲实验室（黄埔） 广州 510700）

摘　要：随着移动通信与人工智能技术的飞速发展，当今社会对于信息智能化处理的需求迅速攀升，信息处理技术正在经历着由信号处理向内容处理的深刻变革。以傅里叶分析为代表的传统信号分析方法在分析和刻画信号时-频特性方面取得了巨大的成功，而在内容层面的信号语义特性分析方法却存在着一定的研究空白。因此，本文致力于为语义信息分析提供新的视角，提出了一种基于语义分解的语义信息分析方法。该方法认为，一个大颗粒度的语义信息可以分解为多个小颗粒度的语义信息，正如傅里叶变换可以将时域信号分解成多个不同频率分量的信号。本文所提出的语义分解方法关注到现实世界中语义信息天然存在的层级结构，其核心思路在于将复杂的语义信息分解为基本的单元组合结构，从而进行表征与理解。为此，本文提出了一系列核心概念，包括语义基元、语义分解树和语义颗粒度等，并定义了语义基元提取的相关运算和方法。通过实验验证，本文提出的语义信息分解方法能够有效地分解和表征语义信号，为信息处理技术的进一步发展提供了理论基础和实践指导。

关键词：语义信息处理　语义分解　人工智能　大模型

中图分类号：TP14

Semantic Information Decomposition Method

Guangming Shi[1,2,3], Yuping Liang[1], Dahua Gao[2,3], Minxi Yang[1], Haixiong Li[1], Xiaodan Song[1]

1 (Xidian University, Xi’an 710071, China)

2 (Pengcheng Laboratory, Shenzhen 518000, China)

3 (Pazhou Laboratory (Huangpu), Guangzhou 510700, China)

Abstract: With the rapid advancement of mobile communication and artificial intelligence technologies, the demand for intelligent information processing in today's society is surging. Information processing technology is undergoing a profound transformation from signal processing to content processing. Traditional signal analysis methods, represented by Fourier analysis, have achieved tremendous success in depicting the physical characteristics of signals. However, there exists a research gap in the analysis of signal semantic characteristics at the content level. This paper aims to provide a new perspective for semantic information analysis by proposing a semantic decomposition-based approach. This method recognizes the inherent hierarchical structure of semantic information in the real world and focuses on decomposing complex semantic information into basic unit combination structures for representation and understanding. The paper introduces a series of core concepts such as semantic primitives, semantic decomposition trees, and semantic granularity, and defines operations and methods related to the extraction of semantic primitives. Experimental validation demonstrates that the proposed semantic information decomposition method can effectively decompose and represent semantic signals, providing a theoretical foundation and practical guidance for the further development of information processing technologies.

Keywords: Semantic Information; Semantic Decomposition; Large Models; Artificial Intelligence

* 通信作者：石光明， gmshi@xidian.edu.cn.

1 引言

在当今社会，信息量的急剧膨胀与用户需求的多样化对信息处理领域提出了双重挑战。同时，人工智能技术正以超乎预期的节奏推动着人类社会快速发展，信息智能化处理与高效率传递的需求在不断攀升，信息处理技术的边界正在被不断拓展[1,2]。

在这一时代背景下，信息处理技术正经历深刻变革，其核心是从传统的信号处理转向更为深入的内容处理。互联网和移动通信技术的广泛应用带来了信息洪流，用户的期望也随之提升，不再满足于简单的信息获取，而是追求高效、有价值的内容消费。这种个性化需求促使信息处理技术不仅实现基础的信号传输，更需提供精准内容服务，语义通信作为下一代通信技术的研究热点应运而生[3-5]。此外，大数据分析、物联网、自动驾驶等新兴产业的崛起，为信息处理技术开辟了新的探索方向，并在公共安全、环境保护、城市规划等领域发挥重要作用，成为经济增长的新动力[6,7]。

传统上，信息处理聚焦于信号的时-频或时-空特性，由此催生了傅里叶分析等经典理论。傅里叶分析通过将复杂的信号分解为简单波形的叠加，揭示了信号的频率成分，广泛应用于信号去噪、频谱分析、系统设计等多个方面[8,9]。然而，随着信息处理要求的提高，当前需求已转向对信息内容的深度分析和理解。尽管信号时-空特性分析方法已经成熟，但在内容理解领域，尤其是语义信息处理方面，仍缺乏系统的理论框架指导。

在信息处理技术向内容处理转型的过程中，信息内容的多样性和复杂性成为显著特征。信息并非孤立存在，而是通过各种结构和关系相互交织。例如，在文本、图像或视频等形式的信息中，不同元素之间存在着显性或隐性的语义关联。因此，理解信息不仅要捕捉单个元素的特性，还需解析它们之间的语义结构。信息的层级结构作为表达这些关联的一种代表性结构，在谱系、空间、时序等多个维度上广泛存在，反映了信息从宏观到微观、从整体到部分的自然分层，有助于更好地把握信息的全貌及其内部逻辑。

类比于傅里叶变换在信号分析中的应用，面向信息内容处理时，提取并刻画语义信息的基本表达单元成为解读其内容的关键切入点。正如傅里叶变换将时域信号分解成多个频率分量一样，语义信息分解旨在识别和分离出构成复杂语义信息的基础单元及其组织结构。这些基础单元不仅承载特定的语义含义，还保留了它们在更大语境中的角色和功能。通过这种方式，我们可以深入挖掘信息的内在逻辑和关联性，从而实现对信息内容的精准理解。具体来说，基础表达单元的提取与刻画具有多重实际应用价值：

1) 首先，它有助于标准化语义信息处理规范，为不同的应用场景提供统一的分析工具和方法论；
2) 其次，通过对语义基元的准确描述，可以深入挖掘信息的内在逻辑和关联性，使智能体或者非专业用户也能快速获取有价值的内容；
3) 此外，基础单元的定义和操作还可以辅助智能系统模拟人脑的工作方式，通过组合和演绎这些基本单元以实现知识发现与推理。

目前，人工智能技术的飞速发展为该方向提供了新的愿景，深度神经网络凭借其高维特征提取与表征能力，在自然语言处理和计算机视觉等诸多现实任务中取得了革命性的成果，大语言模型与大视觉模型[10~16]的涌现更是为赋予了机器理解和刻画世界的可能。这些方法通过大量数据的训练与学习，自动从数据中学习到各类模态信号的显著性特征，以及在不同场

景下的变化规律，似乎能够实现对信号内容的理解。然而，这种由数据驱动的人工智能表征方法往往只聚焦于数据模式与规律的抽取，其致力于从大量数据中捕捉统计上的相关性，而往往难以深入理解数据的深层语义含义，并且忽略语义信息背后的结构关系，型模型中出现的“知识幻觉”现象即为此例证。

综上所述，基于现有信号分析理论的局限和语义层级结构的观察，本文致力于探索一种有效的语义信息分析方法。在作者之前关于语义数学刻画的工作基础上[17,18]，我们构建了新的思路，即通过将复杂的语义信息分解为基本的单元组合结构，为语义信号的深入分析提供新视角。本研究的主要创新点包括：

1) 针对现有语义信息处理方法的局限性，本研究提出了“语义信息分解”的概念，该方法关注现实世界中语义信息的层级结构，定义了“语义基元”、“语义分解树”和“语义粒度”等概念，为语义信息分析提供了新的视角和理论基础；
2) 提出了语义基元的提取方法，利用人工智能大模型的技术优势，在多种应用场景下提取语义基元，实现语义信息的结构化表示；
3) 在公开数据集上通过实验验证了所提出理论的有效性，结果表明本文所提方法能够有效地分解和表征语义信号，为语义信息处理技术的进一步发展提供了实践参考。

2 语义信息分析相关概念

在智能时代，信息处理的需求发生着深刻的转变，为了更好的探讨本文的语义信息分解方法，在本节中将延续文献[17,18]中的讨论，对语义信息分析中的基础概念——语义、信号和信息做出进一步阐述与说明。

首先，有必要明确语义信号的数学刻画。语义作为人脑对外界事物状态的反映，语义信号可以围绕大脑这一中心被区分为物景信号与像景信号两种类型。物景信号指的是那些源自外部世界、不依赖于人类意识，通过传感器探测而得的客观事物状态。例如，光波成像、声波、热量及气味等信号皆属于此类，它们是客观事物时空状态在传感器上的直接投影，不依赖于任何主观意识。物景信号可以通过连续的特征函数 $f(x,t)\in\mathbb{R}^d$ 表示，其中 x 表示事物的空间状态，t 表示事物的时间状态，，该函数描述了客观事物的物理特性而不涉及任何主观解释。相对地，像景信号则是经过大脑加工后产生的语义信号，包括语音、文字等形式，这些信号表征了人脑内部对事物状态的理解与构造 *。像景信号可以用符号化的特征函数 $g(y,k)\in\mathbb{R}^d$ 表示，其中 y 表示语义标签，k 表示该信号的序号。值得注意的是，物景信号与像景信号并非孤立存在，二者之间存在着由相互置信度描述的对应关系。这种关系体现了从物理现实到认知理解的转换过程，即从原始感知数据到具有意义的信息的转变。

信息则是指用户能够理解并从中提取价值的那一部分信号。在语义信息分析领域，用户的关注点在于如何从海量数据中提炼出有意义、可理解的内容，并对其进行后续处理和应用。这不仅限于自然语言处理中的句子或文档含义理解、关系抽取以及情感分析，还包括图像处理中有关纹理识别、目标分类、实例分割等方面的解析。

那么，所谓的语义就是基于群体广泛共识，用某个像景特征信号标注了某个物景特征信号，即实现了两者之间的对齐，从而赋予信号以明确的意义。简而言之，语义是那些已经被

* 在作者已发的文章中，物景和像景信号被分别称为实景和虚景信号。

物景特征或像景特征所标注，并获得广泛认同的特征信号。然而，并非所有的特征信号都能成为语义的一部分，因为某些信号可能尚未完成标注过程或者没有得到足够的群体认同。语义的核心在于表达信号内容。在智能时代，其内涵得到了扩展，涵盖了信息内在逻辑、关联性和意图的表征。这意味着语义不仅仅是简单的数据解读，而是要求智能系统能够模拟人类的认知过程，实现对信息内容的深层次理解和有效利用，进而支持更加智能的决策制定。

3 本文方法

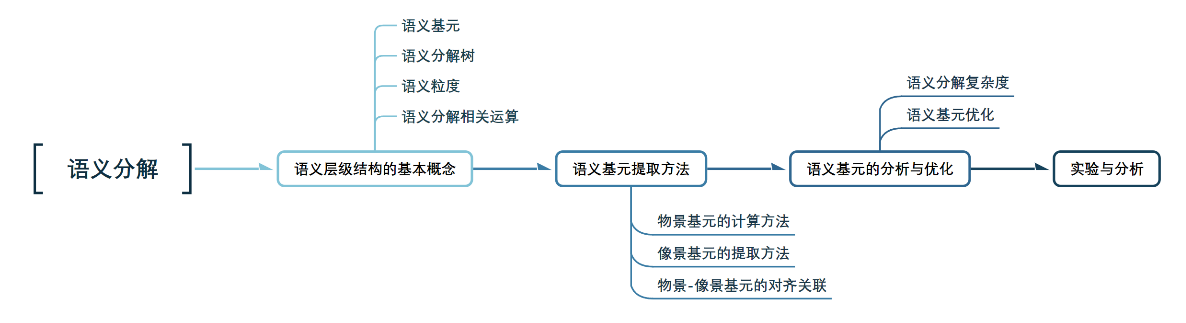


图 1 本文研究思路

针对传统信号分析方法与现有人工智能模型在信息内容理解方面的局限性，本文构建了一种基于语义信息分解的信号分析方法。该方法重点关注现实世界中语义信息普遍存在的层级结构，并提出将复杂语义信息分解为基本单元组合结构的核心思路，从而实现语义信号的表征与理解。本文的研究思路图 1 所示，首先，介绍了语义信号客观存在的层级结构，并定义了“语义基元”、“语义分解树”和“语义粒度”等核心概念。然后，构建了多种场景下的语义基元提取框架，包括语义约束条件下物景基元的计算方法，以及像景基元的提取方法。此外，本文还提出了物景-像景信号的对齐关联方法，尽可能全面、准确地表征语义信息。最后，通过定义语义特征表征复杂度、语义结构表征复杂度和语义分解误差三个指标来评估语义分解方法的效率，并讨论了语义基元信息分解的优化方向。

3.1 语义层级结构的基本概念

语义信息层级结构是客观世界中普遍存在的信息结构，这体现在其对不同领域、不同类型数据的高度概括和统一描述能力。从自然语言处理到图像识别，从生物信息学到社会科学研究，语义信息层级结构均能提供一种有效的组织框架，使得复杂多样的信息得以按照其内在逻辑和关联性进行有序排列。

定义 1 对于物景信号，在其语义信息的层级结构中，语义基元是一类特殊的语义特征函数，其定义了能够独立表达一类语义信息的基础单元，用于描述具有同一簇语义的信号特征函数。具体地，物景信号的语义基元可被表达如下：

$$f = w_1 f_1' + w_2 f_2' + \ldots + w_n f_n' = \sum_{i=1}^{n} w_i f_i' \tag{1}$$

其中，$f_i' \in \mathbb{R}^d$ 表示物景特征函数 f 的语义基元，w_i 表示语义分量，代表基元 f_i 在该分解关系中的语义贡献度。对于物景信号特征函数，由于其客观描述了物体的外部状态，因此可以在数值空间中被分解与合成。

对于像景函数，其描述了客观事物时空状态或者脑中构造出的事物，是离散的语义概念

符号，因此，像景信号的语义基元可被表达如下：

$$g = w_1 g_1' \cap w_2 g_2' \cap \ldots \cap w_n g_n' = \bigcap_{i=1}^{n} w_i g_i' \tag{2}$$

其中，$g_i' \in \mathbb{R}^d$ 表示物景特征函数 g 的语义基元。由于像景函数是人脑中构建出的特定概念，难以通过数值计算得到，因此通常需要在逻辑进行分解与表达。

基于以上定义，语义基元作为表征语义信息的基础单元，具有以下几个关键特征：

1） 独立性：每个语义基元能够独立表达一定的语义信息，这种独立性使得语义基元可以被智能体单独处理和传输。

2） 可操作性：语义基元在系统中可以被操作，如提取、优化、分解和合成等，这种操作性使得系统能够灵活地处理和传输语义信息。

3） 层次性：语义基元之间存在层次关系，并且可以基于语义基元来构建这种层级结构，这种层次性有助于系统理解和处理复杂的语义信息。

4） 可扩展性：开放系统中的语义基元提取通常依赖于先进的深度学习技术，使得系统能够随着技术的发展以及物理环境和任务需求的变化不断扩展和优化基元。

定义 2 直观地，在定义了语义层级结构的基本组成单元后，完整的语义信息层级结构可以由树结构 $T(G,W)$ 来表示，其中语义基元的集合 $G=\{g_1,\ldots,g_n\}$ 构成了 T 的结点集合（此处以像景基元为例），$W=\{w_{ij}\in[0,1] \mid 0<i,j<n\}$ 表示树 T 的边的集合，即语义基元之间的分解关系，边 w_{ij} 的值代表基元 g_j 在基元 g_i 的分解中包含的语义分量，这样的一棵树被称为**语义分解树**。

定义 3 为了反映特定语义基元的微观程度，从而确定其在语义通信系统中的语义层级，可以依据语义基元 v_i 在语义分解树中的深度来定义**语义粒度**（Grained，grd）的概念。具体的：

$$grd(g_i) = \begin{cases} 1 & ,k^{in}(g_i)=0 \\ 1+grd\left(g_j \middle| w_{ji}>0\right) & ,k^{in}(g_i)=1 \end{cases} \tag{3}$$

其中，$k^{in}(g_i)$ 为节点 g_i 入度。值得注意的是，基于树结构的性质，除根节点外的每个叶子结点有且只有一个父节点，因此对于 g_i 至多存在一个 w_{ji} 大于零。由此，(3)式使用了递归的方式表示了节点 g_i 的深度，从而定义其对应基元的语义粒度。图 2 展示了一颗简单的语义分解树及其基元所对应的语义粒度，基元越接近根节点，语义粒度越小，该基元的语义层级越高，其所表示的概念越宏观。

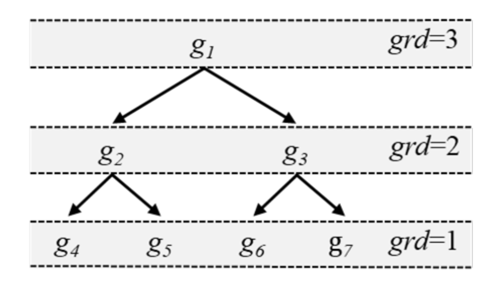


图 2 语义分解树以及语义粒度

在构建基本的像景基元结构表征方式后，为了能够对语义基元以及语义分解树进行有效的操作和分析，接下来通过定义一系列语义分解的相关运算，以实现对语义信号合并、筛选以及修改操作，为后续的基元提取和分析提供基础工具。

定义 4 若语义分解树$T_A(G_A,W_A)$的根节点$g_{A,0}$存在于语义分解树$T_B(G_B,W_B)$的节点集合G_B，则可定义语义分解树的**并运算**，即：

$$T_B(G_B,W_B)\cup T_A(G_A,W_A)=T_{A\cup B}(G_B\cup G_A,norm(W_B\vee W_A)) \tag{4}$$

其中，$\vee$表示取极大值运算，$norm$表示逐粒度的语义分量归一化运算。

定义 5 若语义分解树$T_A(G_A,W_A)$与语义分解树$T_B(G_B,W_B)$的节点集合G_A与G_B的交集不为空，则可定义语义分解树的**交运算**，即：

$$T_B(G_B,W_B)\cap T_A(G_A,W_A)=T_{B\cap A}(G_B\cap G_A,norm(W_B\wedge W_A)) \tag{5}$$

其中，$\wedge$表示取极小值运算。

定义 6 根据语义分解树$T_A(G_A,W_A)$与常数λ，定义其在λ上的**截运算**：

$$T_{A|\lambda}=T_A(\{g_i|w_{\cdot i}>\lambda\},norm\{w_{\cdot i}|w_{\cdot i}>\lambda\}) \tag{6}$$

其中，$w_{\cdot i}$表示语义基元g_i在其父节点中的语义分量。截运算只保留原树中语义分量大于λ的边及其相关的节点。

定义 7 基于语义分解树$T(G,W)$与基元的语义粒度表征，可以定义语义基元在指定语义粒度上的**展开运算**，将宏观的语义信号展开为微观的语义基元集合。语义展开运算的定义如下：

$$\mathcal{D}\left[g_i^{(l)}|L\right]=\begin{cases}\sum_{j=i}^{n}\mathcal{D}\left[w_{ij}\cdot g_j^{(l+1)}|L\right] & ,l<L \text{ and } k^{out}\left(g_i^{(l)}\right)\neq 0\\ w_{ij}\cdot g_i^{(l)} & ,else\end{cases} \tag{7}$$

其中，$\mathcal{D}$表示语义展开函数，$g_i^{(l)}$表示第i个语义基元，其语义粒度为l，L表示展开粒度，$k^{out}\left(g_i^{(l)}\right)$为节点$g_i^{(l)}$出度。

定义 8 基于语义分解树$T(G,W)$与基元的语义粒度表征，便可对于语义基元进行合成运算，从而利用多个细粒度的语义基元来表征更粗粒度的语义基元。其定义如下：

$$\mathcal{D}^{-1}\left[\sum_{j=1}^{n}w_{ij}\cdot g_j^{(l)}\right]=g_i^{(l-1)} \tag{8}$$

语义合成运算可以被视为语义展开运算的逆运算。

3.2 语义基元提取方法

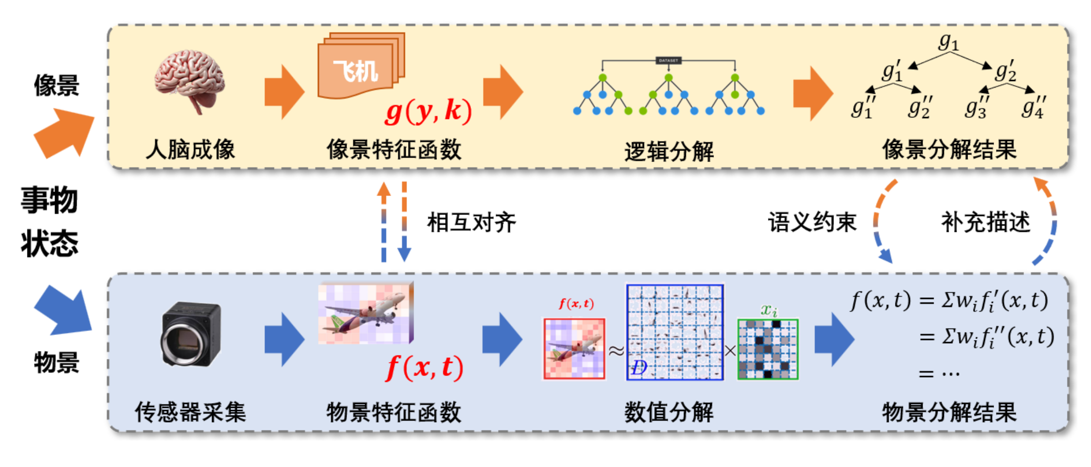


图 3 语义基元提取方法

在构建语义基元的定义之后，其获取方式便是语义信息分解框架中的核心研究内容。如图 3 所示，在面向内容的信息处理框架中，依据处理载体的不同，事物状态可分别由传感器采集与人脑成像所反映，并被描述为物景特征函数与像景特征函数。

物景特征函数是对事物客观状态的数学描述，它是可分析、可计算的。这意味着我们可以通过数值计算方法对物景特征函数进行分析，从而揭示事物的客观状态。例如，在图像处理中，我们可以通过分析图像的像素值、颜色、纹理等特征来描述图像的客观状态。物景语义基元是对物景特征函数的进一步抽象和概括，它是对事物状态层级化的描述。因此，通过基于语义特征函数的矩阵分解方法，能够从物景特征函数中提取出物景语义基元，从而实现对事物状态的更深入理解。

像景特征函数则是对事物状态在人脑中的成像的描述，它是离散的、概念性的。像景特征函数通常通过概念性的符号进行表示，例如语言中的词汇等。像景基元是对像景特征函数的进一步抽象和概括，它是对事物状态在人脑中的成像的更高层次描述。由于像景特征函数是离散的、概念性的，因此像景基元的提取往往依赖于知识表征与逻辑分解。

此外，两种语义模态的分解结果又可以相互对齐，像景基元分解结果可以为物景分解过程提供语义标注与关系约束，而物景基元分解结果也可以辅助挖掘与补充像景基元分解过程中所缺失的概念符号，或者对于场景中冗余的像景基元进行剪枝。总的来说，物景特征函数和像景特征函数分别从客观和主观两个角度描述了事物的状态。通过对于这两类特征函数的分解与分析，我们可以从不同的角度理解事物的状态，从而为面向内容的信息处理提供更全面、更深入的理解。

3.2.1 物景信号的基元提取计算方法

物景信号由传感器采集，能够反应事物的客观时空状态，可以由物景信号特征函数表示 $f(x,t)\in\mathbb{R}^d$，其中 x 表示事物的空间状态，t 表示事物的时间状态。物景语义基元 $f'(x,t)\in\mathbb{R}^d$ 则代表了场景中的多个实体状态 $\{f_1,f_2,...,f_n\}$ 的基本组成单元，例如物体的部件、纹理。基于这样的考虑，在约束条件下对于物景信号的基元提取可以被抽象为以下数学问题：对于 n 个 d 维向量 $\{f_1,f_2,...,f_n\}$，期望找到一组基向量 $\{f_1',f_2',...,f_k'\}$，使得任意 f_i 都可以通过这些基的线性组合来近似表示。该问题实际上可以被视为一个矩阵分解问题，本文提供了一种基于非正交的矩阵分解方法的基础解决思路。

首先，为了从这些综合分析物景信号特征并从中提取出主要的基元，需要一个统一的矩阵来表示这些特征。因此，首先将系统中的语义特征函数 $\{f_1,...,f_n\}$ 堆叠为 $n\times d$ 维的特征矩阵 F。

在现实应用场景中，系统中的语义基元 $\{f_1,f_2,...,f_n\}$ 或许存在着各类天然的语义关系，例如飞机由机身、机翼与尾翼组成，鼠标、键盘、显示器往往出现在同一场景。在求解过程中，将这些语义关系作为基元提取的约束条件，可以确保语义基元提取与语义结构构建的一致性，能够使得提取到的语义基元更加准确的描述显示场景。

具体地，假设存在语义基元 f_1,f_2,f_3 之间存在以下层级语义关系 q_1：

$$q_1: f_1 = 0.6f_2 + 0.4f_3 \tag{9}$$

通过移项并整理为矩阵形式：

$$\begin{bmatrix}-1 & 0.6 & 0.4\end{bmatrix}\begin{bmatrix}f_1\\f_2\\f_3\end{bmatrix}=0 \tag{10}$$

其中，$\begin{bmatrix}-1 & 0.6 & 0.4\end{bmatrix}$为系数矩阵，$\begin{bmatrix}f_1 & f_2 & f_3\end{bmatrix}^T$为特征矩阵。那么，对于整个系统而言，将每组语义关系中的语义变量补全（对于没有涉及的语义变量系数置为0），并全部语义关系组合起来，便可得到该系统的语义关系矩阵Q。基于特征矩阵F与语义关系矩阵Q可构建增广矩阵$[F|Q]$：

$$[F|Q]=\begin{bmatrix}F\\Q\end{bmatrix} \tag{11}$$

该增广矩阵结合了原始矩阵特征函数F和约束条件Q。接下来，即参照式（2）、（3）分别求解协方差矩阵$[F|Q][F|Q]^T$与$[F|Q]^T[F|Q]$的特征值与特征向量，并进一步完成矩阵$[F|Q]$的奇异值分解，使得：

$$[F|Q]=U\Sigma V^T \tag{12}$$

其中，U包含了矩阵$[F|Q]$的左奇异向量，这些向量是$[F|Q]$的行空间的一组正交基，能够描述行向量f_i之间的关系。类似地，V包含了矩阵F的右奇异向量，这些向量是$[F|Q]$的列空间的一组正交基，反映了数据在列空间中的结构，表示了物景特征函数的不同特征或属性之间的相关性。矩阵Σ包含了矩阵$[F|Q]$的奇异值，这些值是矩阵$[F|Q]$的奇异向量对应的标量，它们表示了物景特征函数在这些新坐标轴上的“尺度”。

由此，可以选择前k个最大的奇异值对应的和右奇异向量$\{v_1,\ldots,v_k\}$，作为该系统的语义基元特征函数$\{f_1',\ldots,f_k'\}$，由于在增广矩阵中整合了约束，因此选择的基向量应当自然满足这些约束。接下来，便可使用所提取的语义基元，通过式（5）来描述原始特征函数f_i，此时，式（5）中的a_i由以下方式求解：

$$a_i=V_k^T f_i \tag{13}$$

其中，V_k为向量$\{v_1,\ldots,v_k\}$堆叠而成的矩阵。由此，便在语义先验约束条件下完成物景信号的基元提取。

3.2.2 像景信号的基元提取方法

像景信号是语义概念在人脑中的主观投影，是以语音、语言、文字等抽象符号概念化地表征事物状态。像景信号是构建语义表达的基础单元，类似于字典中的字或词汇。在语义符号系统中，这些基本基元通过特定的规则和结构组合起来，形成更复杂的表达单位，如短语、句子、段落，乃至整个文本或话语体系。语义符号的逐层表达不仅仅是意义的简单叠加，而是一个动态的、结构化的过程，通过不同层次的组合和交互，实现了从基本语义单元到完整概念的有机转变。值得注意的是，语音信号是一类特殊的像景信号，它虽然是连续信号，却是用符号表达语义的信号；图形信号也同样。在分析这类特殊的像景信号时，需要先将其转换成概念符号然后再进行分析处理。

包括语音信号在内的像景信号特征函数可以由$g(y,k)\in\mathbb{R}^d$表示，其中y表示概念的语义标签，k表示该信号的序号。像景语义基元$g'(y,k)\in\mathbb{R}^d$则代表了一组概念$\{g_1,g_2,\ldots,g_n\}$在逻辑

上的基本组成单元。基于这样的考虑，像景信号的基元难以通过计算得到，而需要在知识与经验的引导下，通过逻辑分解获得。因此，为了能够更准确地构建与表述该分解过程，本节将首先定义一系列像景基元的数据结构。

在构建一系列像景基元的数据结构，便可对于像景信号特征函数进行逻辑分解。在封闭环境或者面向特定任务的应用场景下，依据穷举的方式是像景信号分解获取的最简单方法，语义分解树以及语义基元可以依据先验知识一一列举地构建。然而，在更多现实应用场景中，环境可能千变万化，包含着大量的未知和不可预测因素。例如，自动驾驶汽车可能会遇到罕见的天气条件、不寻常的道路状况或新型号的车辆。在这样的应用场景下，语义基元往往不能被一一枚举。针对以上开放场景中的应用限制，人工智能大模型的出现为像景基元的智能提供了可能。大模型通过从海量的数据中学习，能够理解和生成复杂的语义表示，从而更好地适应和理解开放世界中的各种情况。它们不仅可以识别和应对已知的基元，还可以通过泛化能力来处理未见过的基元，这使得它们在处理真实世界的问题时更加灵活和有效。本节提出了一种基础的基于大模型的像景基元构建方法。

算法 1 物景信号的语义分解方法

输入：像景特征函数集合$\{g_1,g_2,...,g_i\}$，视觉大模型 *LVM*，图生成模型 *GCM.*

输出：语义分解树T .

begin

初始化$T(G,W)$， $G\leftarrow\{g_{0,1}\}$， $W\leftarrow\Phi$；

for g_i **in** $\{g_1,g_2,...,g_i\}$：

$G_i\leftarrow LVM(g_i)=\{g'_{i,1},...,g'_{i,k_i}\}$；

$W_i\leftarrow GCM(g'_{i,1},...,g'_{i,k_i})=\{w_{i,j}\in[0,1]\}$；

$T_i(G_i,W_i)\leftarrow G_i,W_i$；

if $g_{i,1}\in G$：

$T(G,W)=T(G,W)\cup T_i(G_i,W_i)$；

else:

$G=G\cup\{g_{i,1}\}$

$W=norm\left(W\cup\{w_{(0,1),(i,1)}\}\right)$

$T(G,W)=T(G,W)\cup T_i(G_i,W_i)$

end for

end

首先，构建一颗只有根节点的初始化语义分解树$T(G,W)$，其中$G=\{g_{0,1}\}$， $g_{0,1}$为系统初始化根节点的基元，$W=\Phi$。其次，基于语言大模型（Large Language Model, LLM）提取像景概念g_i的像景基元集合$LVM(g_i)=\{g'_{i,1},...,g'_{i,k_i}\}$，并且基于图生成模型（Graph Generative Models, GCM）构建基元间的连接关系$W_i=\{w_{i,j}\in[0,1]\,|\,0<i,j<n\}$，便可构建该信号的语义分解树$T_i$。然后，若$g_{i,1}$存在于$G$，则对$T$与$T_i$进行并运算；若不存在，则在$T$中构建一条边，并为

$g_{0,1}$添加一叶子结点$g_{i,1}$，然后再进行T与T_i的并运算。由此，在遍历所有像景信号g_i后，便可得到该系统完整的语义分解树T。上述流程如算法 1 所示。

3.2.3 物景-像景基元的对齐关联

像景与物景基元分别代表了人类对事物状态的主观认知和客观观测，这两种认知方式在本质上是相互关联和相互补充的。物景基元来源于对事物外部特征的客观感知，像景基元则反映了人类对事物的主观认知。尽管物景基元和像景基元在表现形式上存在差异，但它们都指向同一事物或现象的不同方面。例如，一个红色的苹果，其物景基元可能包括其颜色（红色）、形状（圆形）、大小（直径几厘米）等物理特征；而其像景基元可能包括“苹果”这个词汇、苹果的图像、以及与之相关的文化符号等。

人类大脑能够将这两种基元联系起来，使得我们能够通过物景基元来触发或联想到相应的像景基元。这种联系是通过长期的认知和学习过程建立起来的，它使得我们能够从客观的事物特征中抽象出概念，并用语言和图像等符号系统来表达和交流。因此，在完成物景、像景信号基元提取后，进一步地完成物景-像景信号的互补关联具有多重意义。

得益于视觉语言大模型完成（Vision-Language Models, LLM）的发展，这一过程可以通过模型映射得到物景基元所对应的像景基元来完成，即：

$$g_i' = VLM\left(f_i'\right) \tag{14}$$

进一步，便可得到**语义关联对**(f_i', g_i')，用来描述同一客观事物在物景、像景特征上的关联。

在语义信息分解过程中，语义关联对(f_i', g_i')扮演着多重角色，它们强化了物景和像景信号基元之间的互补性，并在提升整个系统的语义表达和处理效率方面发挥着关键作用。具体来说，在像景特征函数对物景基元的标注与约束方面，通过语义分解树和语义展开合成运算，可以形成详细的语义关系表达式，为物景信号的基元计算提供逻辑上的约束，确保计算结果符合预设的语义规则。此外，基于语义关联对，可以对未明确标注的物景基元进行精确的标注，从而优化物景基元的计算过程，使其更符合实际场景的语义结构。

物景基元对像景基元的补充与剪枝方面，利用物景基元的特征函数，可以对像景概念所对应的事物特征进行更加具体和形象的描述，这有助于提升像景基元对事物特征的刻画程度。同时，物景基元可以对像景基元分解结果中存在的冗余基元进行识别和剪枝，显著提升系统在信息表征方面的效率，避免不必要的计算资源浪费。

3.3 语义基元的分析与优化

在构建语义基元提取方法后，本小节深入分析了如何评估所提取的语义基元，特别是针对它们之间的潜在重叠和冗余，提出了优化策略以减少这种冗余并提高表达效率。通过这样的评价分析与进一步优化，进一步提升基元的质量、精炼算法性能，并增强系统的适应性和泛化能力。此外，我们也希望这些讨论能够为后续研究提供有价值的参考方向。

（1） 语义分解复杂度

定义语义分解评估指标是分析语义信息分解方法效率的首要前提，本节定义了三个重要的指标：特征表征复杂度、结构表征复杂度和语义分解误差。这三个指标分别从不同角度衡量了语义分解的有效性、效率和准确性，为语义信号处理和分析提供了重要的参考依据。

定义 9 系统的**特征表征复杂度**Θ_f主要来自于物景特征函数对于事物状态的描述，其可以

由单位物景语义基元 f_i 的表征复杂度 θ 与特征函数数目 $n+k$（n 为原始特征函数数目，k 为基元特征函数数目）表示：

$$\Theta_f = (n+k)\theta \tag{15}$$

定义 10 系统的**结构表征复杂度** Θ_g 主要来自于像景语义分解树结构的构建，可以由全体像景基元的平均查找复杂度表示，而语义基元 g_i 在语义分解树中的查找复杂度与其在树中的深度相关，也就可以由其语义颗粒度表示，因此可定义：

$$\Theta_g = \frac{1}{k}\cdot\sum_{i=1}^{k} grd(g_i') \tag{16}$$

定义 11 系统的**语义分解误差** E 用来描述语义分解结果的准确性和完整性，可由原始特征函数分解误差 ε_i（式（5））的平均值得到：

$$\mathrm{E} = \frac{1}{n}\cdot\sum_{i=1}^{n}\|\varepsilon_i\|_1 = \frac{1}{n}\cdot\sum_{i=1}^{n}\left\|\sum_{j=1}^{k} w_j f_i' - f_i\right\|_1 \tag{17}$$

（2） 语义基元的优化

面向内容处理的语义信息分解任务存在一个直观目标，即使得语义分解复杂度 Θ_f、Θ_g 在应用场景允许的情况下，使得语义分解误差 E 尽可能地小，由此，这样的问题可以表述为离散的优化问题。将式(17)~(19)带入，整理可得优化目标函数：

$$\begin{aligned}&\min_{f_i', g_i', k}\frac{1}{n}\cdot\sum_{i=1}^{n}\left\|\sum_{j=1}^{k} w_j f_i' - f_i\right\|_1\\ &s.t.\ (n+k)\theta \le \Theta_f^{\max}\\ &\quad \frac{1}{k}\cdot\sum_{i=1}^{k} grd(g_i') \le \Theta_g^{\max}\end{aligned} \tag{18}$$

其中，$\Theta_v^{\max}$ 与 $\Theta_v^{\max}$ 分别为通信系统的语义特征复杂度上限与语义结构复杂度上界。进一步，对于固定的通信系统，原始语义特征函数的数目 n 应当为常数，因为，(21)可进一步化简为：

$$\begin{aligned}&\min_{f_i', g_i', k}\sum_{i=1}^{n}\left\|\sum_{j=1}^{k} w_j f_i' - f_i\right\|_1\\ &s.t.\ k\theta \le \Theta_f^{\max}\\ &\quad \frac{1}{k}\cdot\sum_{i=1}^{k} grd(g_i') \le \Theta_g^{\max}\end{aligned} \tag{19}$$

由此，便将语义基元优化目标函数建模为式（22），以便引导接下来的语义信号分析与优化。

4 实验分析

为验证本文所提出的语义信息分解方法，以图像模态的物景信号为例，本文提取了 COCO 2017 数据集[19]的语义基元，从而验证了该方法在实际应用中的可行性与有效性。

4.1 实验配置

COCO 2017 数据集是一个广泛用于图像识别、分割和标注的大型数据集，包含约 12 万张

图像、150 万个物体实例、80 个物体类别、91 种物品、25 万个图像分割区域。它旨在推动计算机视觉领域的研究，特别是目标检测、分割和场景理解等任务。

本次实验中实景特征函数的提取是通过 VLPart[20]实现的。VLPart 是一个基于 Mask R-CNN 架构的视觉-语言集成模型，其分类器利用 CLIP[21]中的类别名称文本嵌入。该模型能够在部件级、对象级和图像级数据上无缝训练检测器，进一步分析图像数据到其组成部分，并在各种部件分割数据集上达到了最先进的性能。因此，VLPart 提取的部件可以作为图像中的实景语义基元。

在实验环境方面，实验基于 Python 3.9 和 PyTorch 1.12.1 实现，并在配备了单个 RTX 4090 GPU 和 Intel i7-5930K CPU 的 Windows 11 系统上进行。

4.2 结果分析

图 4 展示了基元提取的可视化结果，可以看出 VLPart 模型能够有效地识别图像中的物体部件，并将其作为语义基元。例如，在“table”图像中，模型识别出了桌面、桌腿和桌面上的物品等部件，在“dog”图像中，模型识别出了狗的头部、身体和四肢等部件，在“truck”图像中，模型识别出了车轮、车身和车头等部件。

图 5 展示了语义分解树的可视化结果，可以看出每个图像都可以被分解为多个语义基元，并通过树形结构表示其层级关系。例如，在“table”图像中，桌面是根节点，桌腿和桌面上的物品是子节点，在“dog”图像中，头部是根节点，身体和四肢是子节点，在“truck”图像中，车身是根节点，车轮和车头是子节点。

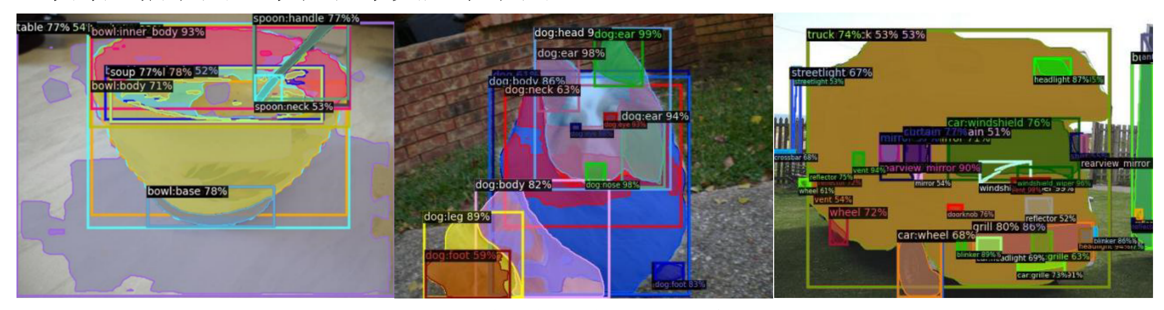
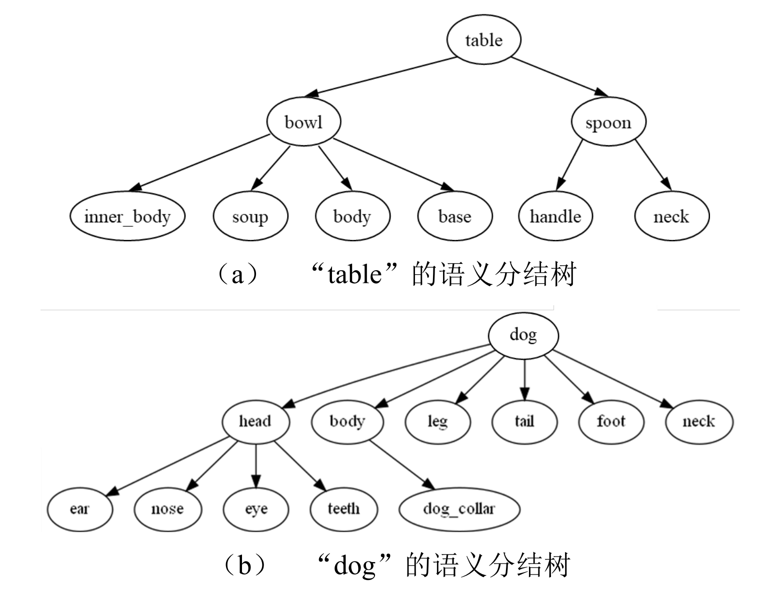
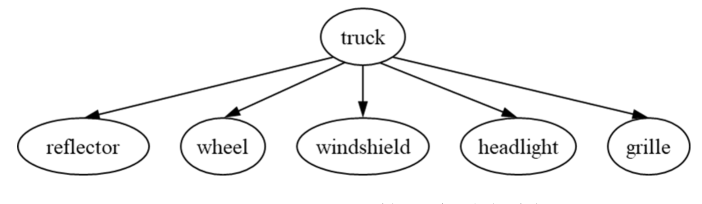


图 4 基元提取的可视化结果



（a） “table”的语义分结树

（b） “dog”的语义分结树



（c）　“truck”的语义分解树

图 5 语义分结树的可视化结果

以上实验结果表明，本文提出的语义信息分解方法能够有效地将图像信息分解为具有不同表示形式的语义特征，并通过树形结构表示其层级关系。这为后续的语义信号分析和处理提供了基础，并为构建更加智能和高效的信息处理系统提供了有力支撑。

5 结论

面向当今社会对于信息处理技术的新需求，本文致力于探索一种有效的语义信息分析方法。具体地，本文构建了语义信息分解的理论框架，定义了语义基元、语义分解树和语义粒度等核心概念，并提出了语义分解运算，为语义信息分析提供了新的视角和理论基础；提出了语义基元的构建方法，针对封闭系统和开放系统两种情况，分别提出了线性基元构建方法和非线性基元构建方法，为语义基元的获取提供了技术支撑；验证了语义分解方法的有效性，通过实验验证了语义分解方法能够有效地分解和表征语义信号，为语义信息处理技术的进一步发展提供了实践参考。未来，该工作将继续深入研究语义分解理论，完善更多场景中的语义基元提取方法，并进一步构建语义分解方法评价体系，期望为构建更加智能和高效的信息处理系统提供有力支撑。

参考文献：

[1] 孙学胜.智能信息处理技术的发展和应用研究[J].信息记录材料, 2023, 24(4):119-121.

[2] 王静静,叶鹰,王婉茹.GPT 类技术应用开启智能信息处理之颠覆性变革[J].图书馆杂志,2023,42(5):9-13.

[3] 石光明，肖泳，李莹玉，等．面向万物智联的语义通信网络[J]. 物联网学报，2021.DOI:10.11959/j.issn.2096-3750.2021.00209.

[4] Shi G , Xiao Y , Li Y ,et al.From Semantic Communication to Semantic-Aware Networking: Model, Architecture, and Open Problems[J].IEEE Communications Magazine, 2021, 59(8):44-50.DOI:10.1109/MCOM.001.2001239.

[5] Xiao Y , Sun Z , Shi G ,et al.Imitation Learning-Based Implicit Semantic-Aware Communication Networks: Multi-Layer Representation and Collaborative Reasoning[J].IEEE Journal on Selected Areas in Communications, 2022, 41:639-658.DOI:10.1109/JSAC.2022.3229419.

[6] 余文科，程媛，李芳，等．物联网技术发展分析与建议[J]. 物联网学报，2020(004):004.DOI:10.11959/j.issn.2096-3750.2020.00195.

[7] 章军辉，陈大鹏，李庆．自动驾驶技术研究现状及发展趋势[J]. 科学技术与工程， 2020, 20(9):10.DOI:CNKI:SUN:KXJS.0.2020-09-005.

[8] 杨丽娟，张白桦，叶旭桢．快速傅里叶变换 FFT 及其应用[J]. 光电工程， 2004, 31(S1):1-3.DOI:10.3969/j.issn.1003-501X.2004.z1.001.

[9] 何勇福，潘芳芳．信号的傅里叶级数展开及其应用[J]. 教育界：高等教育，2013(9):2.DOI:10.3969/j.issn.1674-9510.2013.27.064.

[10] Bubeck S, Chandrasekaran V, Eldan R, et al. Sparks of artificial general intelligence: Early experiments

with gpt-4[J]. arXiv preprint arXiv:2303.12712, 2023.

[11] Thoppilan R, De Freitas D, Hall J, et al. Lamda: Language models for dialog applications[J]. arXiv preprint arXiv:2201.08239, 2022.

[12] GLM T, Zeng A, Xu B, et al. ChatGLM: A Family of Large Language Models from GLM-130B to GLM-4 All Tools[J]. arXiv preprint arXiv:2406.12793, 2024.

[13] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[14] Kirillov A, Mintun E, Ravi N, et al. Segment anything[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 4015-4026.

[15] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//International conference on machine learning. PMLR, 2021: 8748-8763.

[16] Liu H, Li C, Wu Q, et al. Visual instruction tuning[J]. Advances in neural information processing systems, 2024, 36.

[17] 石光明，高大化，杨旻曦，等．信号的语义刻画与度量[J]. 电子学报， 2022, 50(9):2068-2078.DOI:10.12263/DZXB.20210238.

[18] 石光明，杨旻曦，高大化，等．面向语义信息直传的通信架构[J]. 通信学报,2023,44(5):15-27. DOI:10.11959/j.issn.1000−436x.2023098.

[19] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.

[20] Sun P, Chen S, Zhu C, et al. Going denser with open-vocabulary part segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 15453-15465.

[21] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//International conference on machine learning. PMLR, 2021: 8748-8763.